The Design of the Data Processing Project for China's Fifth National Population Census

Zhang Shouqing
No75, Yuetannanjie, Sanlihe, Beijing P.R.C
zsq@satats.gov.cn

The work of data processing of the China' fifth national population census with computers is a huge systematic engineering. To adopt a model of data processing suited Chinese situation and scientifically design the data processing project is important guarantee to fulfill the task of data processing.

1.**The basis of design for the data processing project**

(1) From the point of view of development of China' s productive forces, It is possible to adopt the modern information technology (IT) for the population census.

(2) From the point of view of four management stages of state, province, prefecture and county levels at statistical field, to adopt a work model of "data entry by OCR, image storage, stetting up net stations, and data processing at three stages" for China' s fifth national population census is to ensure the implementation of the China' s fifth national population census smoothly in the light of organizational work, that is to carry out data entry by OCR at prefectures, to set up net stations at state and provinces, to fulfill data processing at state, provinces and prefectures.

(3) From the point of view of computer application in the statistical field, first, we have used OCR technology for the work of data processing of project of China' s First National Agricultural Census in 1997. Second, to store census forms and images has characteristics of small volume, large capacity, easy keep and operation, and save money. Third, the construction of backbone network of state statistical information engineering can support the on-line data processing of the population census. Fourth, database, meta database have been put into application of statistic system for years and with application of database basis.

(4) From the point of view of need and requirement, volume of data processing, the basic work of the data processing of the China' s fifth national population census is to full use construction result of China' s Statistical Information Engineering , to take the data entry, edit, check, collection and set up data base by application of modern information techno logy, and for some cities with conditions to set up population geographical information system. The total volume of long forms, short forms, death forms and temporarily stay forms for China' s fifth National Population Census is up to 410,000,000 sheets. The volume of initial data entry is up to 64,000,000,000 characters, it is equal to 64GB for computer store. The time for data entry is 10 months.

In order to meet needs of society and governments, we have to take the data processing of with modern information technology such as OCR, image store, Internet/Intranet and database so that we may finish the data entry at period of 10 months for such huge data quantity.

2. **Technical Frame of the Data Processing Project:**

(1) The OCR technology means that the original forms filled by hands are to be scanned into computer through scanners, instead of keying by hands.

(2) The image store technology means that the census forms, images scanned to computers are etched on optical disks, keeping them instead of paper.

(3)The application of Internet/Intranet technology is to fully use various computer sources, set up an easy-operation application system, process and transmit data through on- line network.

(4) The data base and meta base technology means to process, manage data with computers, to manage, maintain data nature, character, classification category and professional regulation with data base technology.

(5) To adopt ECA certificate system could ensure to safety of system operation and data keeping.

3. **Requirement for software and hardware environments**

For the China's fifth national population census, we unified software and hardware environments in order to increase the system efficiency and management.

(1) Requirement for software environment

The equipment requirement for prefecture level is that operation system is Chinese WIN98 (SE). Development tool is Visual Studio 6.0 C++(development environment). Development software are OCR software, edit-check software, prevent virus system for neighbor-net.

The equipment requirement for state and province level is that operation system is HP UNIX; Data base management system is Oracle 8. WebServer is Oracle/Internet Application Server (OAS/IAS); Development tool is HP UNIX C++(only for state level).

The software for data processing is OCR software system, edit-check system software, data processing system software for state and province levels, ECA certificate system and SSL data cipher system.

(2) Requirement for hardware environment

Data entry by OCR equipment system is for prefecture level but Web server and database server is for state and province level.

(3) Requirement for network environment

Data entry by OCR for prefecture level is taken under Ether Net of WIN95/98. Through neighbor-net connected computers with HUB, a close, special, integrated platform for China's fifth national population census is taken shape. Data is up transmitted by professional software through prefectures' WAN connected with HUB. While data processing for China's fifth national population census at state and province level is taken under key net of statistical information system..

4. **Flowchart for data processing is omitted here.**

In one word, the design of the data processing of China's fifth national population census fully appears the modern information technology and technical elements. Which is sure to fulfill with high quality, efficiency as well as to utilize population information source and to train a large patch of computer technicians.